



10

(11) Publication number : 11-015489
(43) Date of publication of application : 22.01.1999
(51) Int.Cl. G10L 3/00 G10H 7/08 G10K 15/04 G10L 9/02
(21) Application number : 09-181816
(71) Applicant : YAMAHA CORP
(22) Date of filing : 24.06.1997
(72) Inventor : OTA SHINICHI; and NISHIMOTO TETSUO
(54) SINGING SOUND SYNTHESIZING DEVICE

(57) Abstract:

PROBLEM TO BE SOLVED: To synthesize a more natural singing sound on the basis of text data.

SOLUTION: Text data are read out corresponding to melody data stored in a music information storage part 5, and voice parameters consisting of phonematic formant data and phonematic articulation coupling control data corresponding to the vocal sound of the text data are read out from a voice quality control information storage part 6. A voicing parameter supply control part 7 interpolates the respective parameters and supplies them to formant waveform generation parts 81 to 8m at specific intervals of time to synthesize and output a singing voice corresponding to the text. The speed of pitch variation at the time of a rise in interval is made slower than that at the time of a decrease. Further, the pitch is held when a voiceless sound is generated and begins to be varied when a voiced sound is generated. For consonant and vowel sounds, the target values of the formant frequencies of the vowels are varied.

END

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-15489

(43) 公開日 平成11年(1999) 1月22日

(51) Int.Cl.⁸
G10L 3/00
G10H 7/08
G10K 15/04
G10L 9/02

識別記号

302

F I

G10L 3/00 J
G10K 15/04 302D
G10L 9/02 L
G10H 7/00 531

審査請求 未請求 請求項の数 3 F D (全 14 頁)

(21) 出願番号 特願平9-181816

(22) 出願日 平成9年(1997) 6月24日

(71) 出願人 000004075

ヤマハ株式会社

静岡県浜松市中沢町10番1号

(72) 発明者 大田 慎一

静岡県浜松市中沢町10番1号 ヤマハ株式会社内

(72) 発明者 西元 哲夫

静岡県浜松市中沢町10番1号 ヤマハ株式会社内

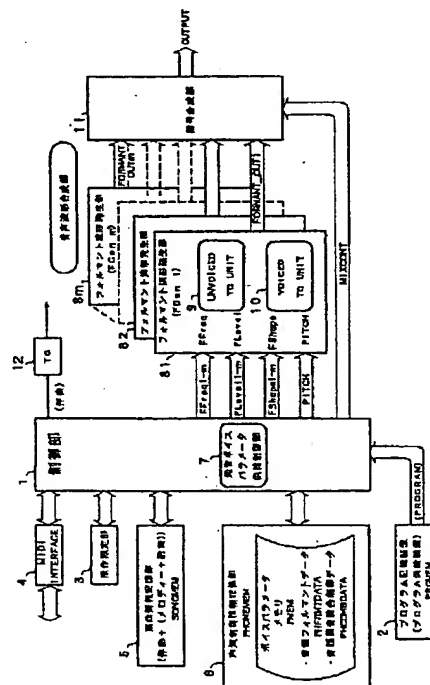
(74) 代理人 弁理士 浅見 保男 (外2名)

(54) 【発明の名称】 歌唱音合成装置

(57) 【要約】

【課題】 歌詞データに基づき、より自然な歌唱音を合成する。

【解決手段】 楽曲情報記憶部5に格納されたメロディーデータに対応して歌詞データを読み出し、声質制御情報記憶部6から前記歌詞データの音韻に対応する音韻フォルマントデータと音韻調音結合制御データとからなるボイスパラメータを読み出し、発音ボイスパラメータ供給制御部により、各パラメータを補間して所定時間毎にフォルマント波形発生部8 1〜8 mに供給して、歌詞に対応する歌唱音を合成出力する。音程の上昇時におけるビッチ変化の速度を下降時の速度よりも遅くする。また、無声音の発声時にはピッチを保持し、有声音の発音時にピッチの変更を開始する。子音母音子音のときは、母音のフォルマント周波数の目標値を変更する。



【特許請求の範囲】

【請求項 1】 歌詞データとメロディデータとを記憶し、該メロディデータの読み出しに対応して前記歌詞データを読み出し、該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成された歌唱音合成装置において、

発音させる音韻のピッチを上昇させるときのピッチ変化速度が、発音させる音韻のピッチを下降させるときのピッチ変化速度よりも遅い速度とされていることを特徴とする歌唱音合成装置。

【請求項 2】 歌詞データとメロディデータとを記憶し、該メロディデータの読み出しに対応して前記歌詞データを読み出し、該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成された歌唱音合成装置において、

有声音、無声音、有声音の順に発音させる場合において発音する音韻のピッチを変化させるときには、前記無声音の発音期間は先行する前記有声音のピッチデータを保持し、後続する前記有声音の発音開始時から当該ピッチの変化を開始するようになされていることを特徴とする歌唱音合成装置。

【請求項 3】 歌詞データとメロディデータとを記憶し、該メロディデータの読み出しに対応して前記歌詞データを読み出し、該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成された歌唱音合成装置において、

子音、母音、子音の順に発音させる場合において前記母音を発音させるときは、そのフォルマント中心周波数が当該母音を単独で発音させる場合のフォルマント中心周波数に達しないうちに、後続する前記子音のフォルマント中心周波数に戻すようになされていることを特徴とする歌唱音合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、歌詞データに基づいて対応する音韻を発音し、当該歌詞を人声音で歌唱するようになされた歌唱音合成装置に関する。

【0002】

【従来の技術】 音声合成技術の一つとして、フォルマント合成による音声合成方式が知られている。この方式は、時系列的に変化するフォルマントに関するパラメータデータを複数ステップにわたって予め記憶している記憶手段と、音声を発声すべきときに前記記憶手段から前記パラメータデータを複数ステップにわたって時系列的に読み出す読出手段と、読み出されたパラメータデータが入力され、該パラメータデータに応じて決定されるフォルマント特性を持つ楽音信号を合成するフォルマント合成手段とを備え、音声信号のフォルマントを時系列的に変化させるものである。

【0003】 最近では、このような音声合成技術を音楽

に適用し、歌詞データに基づいて自然な歌唱音を合成出力する歌唱音合成装置（シンギングシンセサイザ）が提案されている（特開平 9-50287 号公報を参照されたい）。この歌唱音合成装置は、歌詞データとメロディデータとを記憶し、該メロディデータの読出しに対応して前記歌詞データを読み出し、当該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成されている。そして、一つの音符を発音する時間内に複数の音節データを発音することができるようにし、また、子音の発音時間は設定された時間とし、母音の発音時間は音符長により変化させるようにして、自然な歌唱音を合成出力することができるようになされている。

【0004】

【発明が解決しようとする課題】 しかしながら、自然な歌声に聞こえるためには、上記した以外にも歌唱特有の問題がある。例えば、歌唱においては音程が変化するため、ピッチ変化が当然に現われる。図 7 の（a）は、a という音を音程を変えつつ発音したときのピッチ周波数の変化の様子を実測した図である。この図に示すように、低い音程から高い音程に上げたときと高い音程から低い音程へ下げたときでは、ピッチ変化の傾きが異なっている。すなわち、高い音程に上がるときのピッチ変化の傾きは、低い音程に下がる時の傾きよりも緩やかなものとなっている。音程の上下は声帯にある筋肉の緊張、弛緩により行なわれるが、その力学的特性に起因するものと考えられる。

【0005】 従来の歌唱音合成装置においては、図 7 の（b）に示すように、ピッチの変化をピッチの上昇時下降時のいずれにおいても同様に行なっていた。したがって、特にピッチの下降時に自然の発声によるものではない違和感があった。したがって、図 7 の（c）に示すように、ピッチの上昇時と下降時においてピッチ変化の傾きを変更させることが、人間の聴覚上違和感のないものとするために必要となる。

【0006】 また、図 8 は、有声音（voiced sound）V の後にピッチの異なる無声音（unvoiced sound）U と有声音 V からなる音節を発音させるときのピッチの変化の様子を説明する図である。同図（a）に示すように、このような場合には無声音 U の発音開始時からピッチ変化を生じさせるのが通常であるが、無声音 U にはピッチ成分がないため、無声音 U の発音期間においてはピッチ変化が出力されない。したがって、図 8 の（b）に示すように、当該音節の母音 V の発音期間になってから、いきなりピッチが跳んだ母音 V が発生されることとなる。したがって、ピッチのつながりが不連続な不自然な発音になってしまう。そこで、このような不自然さがなく、図 8 の（c）に示すように、自然なピッチのつながりを有する音声出力が望まれる。

【0007】 さらに、実際の人間の発音においては、子音 C（consonant）、母音 V（vowel）、子音 C というよ

うに子音の間に母音が挟まれているときには、その母音の本来のフォルマント中心周波数に達することなく次の子音のフォルマント中心周波数に変化していくこと（いわゆる、アンダーシュート）が発生する。図9はこの様子を示す図である。例えば母音uを単独で発音した場合には、図9の（a）に示すような第1フォルマント中心周波数FF1および第2フォルマント中心周波数FF2となっている。しかしながら、d-u-dと発音した場合には、図9の（b）に示すように、母音uの第2フォルマント中心周波数FF2は、同図（a）に示したFF2の周波数まで達することなく、次のdの第2フォルマント周波数に遷移している。このように、子音で挟まれた母音は、実際には、単独で発音されたときのフォルマント中心周波数とは異なるフォルマント中心周波数で発音されている。

【0008】そこで、本発明は、上述のような実際の人間による発音に現われる特徴を取り込むことにより、より自然な歌声を発声することができる歌唱音合成装置を提供することを目的としている。

【0009】

【課題を解決するための手段】上記目的を達成するために、本発明の歌唱音合成装置は、歌詞データとメロディデータとを記憶し、該メロディデータの読み出しに対応して前記歌詞データを読み出し、該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成された歌唱音合成装置において、発音させる音韻のピッチを上昇させるときのピッチ変化速度が、発音させる音韻のピッチを下降させるときのピッチ変化速度よりも遅い速度とされているものである。これにより、人間による発声の場合と同様の音程の変化を実現でき、自然な歌唱音を発音させることができる。

【0010】また、本発明の他の歌唱音合成装置は、歌詞データとメロディデータとを記憶し、該メロディデータの読み出しに対応して前記歌詞データを読み出し、該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成された歌唱音合成装置において、有声音、無声音、有声音の順に発音させる場合において発音する音韻のピッチを変化させるときには、前記無声音の発音期間は先行する前記有声音のピッチデータを保持し、後続する前記有声音の発音開始時から当該ピッチの変化を開始するようになされているものである。これにより、ピッチのつながりがスムーズになり、自然な発音となる。

【0011】さらに、本発明のさらに他の歌唱音合成装置は、歌詞データとメロディデータとを記憶し、該メロディデータの読み出しに対応して前記歌詞データを読み出し、該歌詞データに対応した音韻を発音させて当該歌詞を歌唱させるように構成された歌唱音合成装置において、子音、母音、子音の順に発音させる場合において前記母音を発音させるときは、そのフォルマント中心周波

数が当該母音を単独で発音させる場合のフォルマント中心周波数に達しないうちに、後続する前記子音のフォルマント中心周波数に戻すようになされているものである。これにより、人間による発生と同様の音韻を発声することができ、自然な歌唱音を合成することができる。

【0012】

【発明の実施の形態】図1は、本発明の歌唱音合成装置のシステム構成の一例を示す図である。この図において、1は装置全体の制御を行う制御部であり、CPUおよび各種フラグや各種変数領域、バッファ等として使用されるRAMなどが設けられている。2は制御プログラムなどが格納されているプログラム記憶装置PRGMEMである。なお、このプログラム記憶装置2は、ネットワーク等を介して制御プログラムを前記制御部1に供給するプログラム供給装置であってもよい。3は機器の動作状態や入力データおよび操作者に対するメッセージなどを表示するための表示部、操作つまみや操作ボタンなどの各種設定操作子などが設けられた操作設定部である。4はMIDIインターフェース部であり、外部のMIDI機器やネットワーク等に接続されており、MIDIデータの入出力を行なう。5はメロディデータおよび歌詞データからなる歌唱データや伴奏データからなる楽曲情報が格納される楽曲情報記憶部SONGMEMであり、半導体メモリのほか、フロッピーディスク装置、ハードディスク装置、MOディスク装置、ICメモリカードなどの各種メディアを使用することができる。

【0013】また、6は声質制御情報記憶部PHONEMEMであり、この声質制御情報記憶部PHONEMEM6内には、音韻フォルマントデータPHFRMTDATAおよび音韻調音結合制御データPHCOMBDATAが格納されているボイスパラメータメモリPMEMが設けられている。このボイスパラメータメモリPMEMは、発生する音声の種類、例えば、男声、女声、あるいは特定の歌手等ごとにそれぞれ設けられており、操作者は、発音させたい声質に応じて使用するボイスパラメータメモリPMEMを選択することができるようになされている。前記音韻フォルマントデータPHFRMTDATAは、各音韻をフォルマント合成するためのパラメータであり、各音韻対応に、その音韻を発生するための各フォルマントの形状を指定するデータ（FShape 1~m）、各フォルマントの中心周波数をそれぞれ指定するデータ（FFreq 1~m）、各フォルマントの出力レベルデータ（FLevel 1~m）等のデータからなっている。また、前記音韻調音結合制御データPHCOMBDATAはある音韻から他の音韻に移行する際の調音結合（特に、フォルマント周波数遷移）を行なうための各種パラメータであり、前記音韻フォルマントデータPHFRMTDATA中の各データに対する補間レートや該補間レートを補正するための定数等からなっている。

【0014】7は発音ボイスパラメータ供給制御部であり、前記楽曲情報記憶部SONGMEM5から当該楽曲のメロ

ディデータおよび歌詞データを読み出し、歌詞データに対応する音韻フォルマントデータPHFRMTDATAおよび音韻調音結合制御データPHCOMBDATAを前記声質制御情報発生部PHONEMEMのボイスパラメータメモリPMEMから読み出して補間演算を実行し、所定時間毎にフォルマント波形発生部81~8mに発音制御パラメータとして供給する。

【0015】81~8mはフォルマント波形発生部FGen 1~mであり、前記制御部1の発音ボイスパラメータ供給制御部7から所定時間毎に供給される発音制御パラメータ、すなわち、フォルマント中心周波数FFreq 1~m、フォルマント振幅レベルFLevel 1~m、フォルマント形状情報FShape 1~mおよび音高情報PITCHに基づいてそれぞれ対応する次数のフォルマント波形を発生するものである。図示するように、各フォルマント波形発生部FGen 1~mには、それぞれ、無声音の発生を受け持つ無声音音源ユニットUNVOICED TG UNIT9と有声音の発生を受け持つ有声音音源ユニットVOICED TG UNIT10とが設けられており、各音源ユニット9および10の出力は加算されて、当該フォルマント波形発生部FGen n (n=1~m)の出力FORMANT_OUT n (n=1~m)として出力される。なお、このような音声合成装置は本出願人により既に提案されている(特開平3-200299号公報)。また、前記各フォルマント波形発生部81~8mは楽音の発生も行うことができ、音声のための発音チャンネルとして割り当てられていない音源は、楽音の発生に割り当てることができる。

【0016】11は信号合成部であり、前記m個のフォルマント波形発生部81~8mから出力される出力信号FORMANT_OUT 1~mを制御部1から供給される制御信号MIXCONTに応じて加算して出力する信号合成部である。

【0017】このように構成された本発明の歌唱音合成装置の動作について説明する。図2は、制御部1において実行されるメインルーチンの動作フローチャートである。さて、動作が開始されると、まず、ステップS10においてシステムの初期化が行われる。続いて、ステップS20の操作イベント検出処理に進み、前記操作設定部3において操作イベントが発生したか否かが判定される。次に、ステップS30に進み、前記操作イベント検出処理S20において検出した操作イベントが演奏する楽曲および歌唱の選択イベントである場合には、対応する楽曲および歌唱に対応するデータ(伴奏データ、メロディーデータおよび歌詞データ)を前記楽曲情報記憶部SONGMEM5から読み出す。なお、このメロディーデータは、MIDIのシーケンスデータと同様のデータとされており、KEYON、KEYOFF、音高[PITCH]、音長[NOTELENGTH]、タッチ[TOUCH]等のデータが含まれている。また、歌詞データは、例えば、音符と歌詞との関連を持たせるために音符の区切りを示す区切り符号が付加された、歌詞に対応する音韻が記載されたデータである。さらに、伴奏データは、MIDIのシーケンスデータとされてい

る。

【0018】続いて、ステップS40に進み、前記ステップS30において読み出された選択された楽曲の伴奏データの演奏処理が行われる。これは、通常の楽音発生処理と同一の処理であり、伴奏用の楽音が前記音源部7により発生される。続いて、ステップS50に進み、当該楽曲の歌唱音制御処理が行われる。この処理により、前記メロディーデータおよび歌詞データに対応した歌唱音が前記フォルマント波形発生部FGen 1~mにおいて発生され、前記信号合成部11において合成されて出力されることとなる。なお、この歌唱音制御処理S50の詳細については後述する。続いて、ステップS60に進み、前記ステップS20において検出された処理が前記楽曲・歌唱選択処理以外の処理である場合におけるその操作に対応する処理や表示制御処理を行う。そして、前記ステップS20に戻り、再び、上述した処理を繰り返し順次実行する。

【0019】次に、前記S50の歌唱音制御処理について説明する。図3は、この歌唱音制御処理の動作フローチャートである。なお、このフローチャートにおいては、説明を簡単にするため、発生する歌唱音の振幅の制御については省略してある。この歌唱音制御処理が起動されると、まず、ステップS301において、歌唱音発生処理中であるか否かが判定される。この歌唱音発生処理中であるか否かは、前記制御部1におけるRAM上に設定されたフラグSINGING_STARTが「1」であるか否かにより判定される。このフラグSINGING_STARTが「0」のときは歌唱音発生処理中でないと判断され、ステップS302に進む。そして、前記操作設定部3において歌唱音発生開始指示イベントが発生したか否かを判定する。その結果、歌唱音発生開始指示イベントが発生していなかったときは、この歌唱音制御処理S50を終了し、前記ステップS60のその他処理に進む。

【0020】また、歌唱音発生開始指示イベントが発生していたときには、前記S302の判定結果がYESとなり、歌唱音発生処理初期化処理S303が行われ、前記歌詞データから読み出す音韻のシーケンスポインタiに初期値「1」がセットされる。次に、ステップS304に進み、前記フラグSINGING_STARTに「1」をセットして、この回の歌唱音制御処理を終了する。

【0021】さて、フラグSINGING_STARTが「1」にセットされている状態(歌唱音発生処理中の状態)で、この歌唱音制御処理S50が開始されたときは、前記ステップS301の判定結果がNOとなりステップS305に進む。そして、このステップS305において、歌唱音発生終了指示イベントが発生したか否かを判定する。この判定の結果がYESのときは、ステップS306において前記フラグSINGING_STARTを「0」にリセットして歌唱音発生処理終了処理を行い、歌唱音制御処理を終了する。

【0022】また、前記ステップS305の判定結果がNOのときは、ステップS307に進み、第*i*番目の音韻を発音を開始すべきタイミングになったか否か、すなわち、前記メロディーデータのノートオンイベントのタイミングとなったか否かを判定する。このステップS307の判定の結果がYESのときは、この音韻*i*を発生するための準備処理が実行される。まず、ステップS308において、メロディーおよび歌詞イベントの分析が行われる。具体的には、前記メロディーデータのノートオンイベントに含まれている音高情報KCPITCH、現在発音中の音韻(*i*-1)、その音高(*i*-1)、その発音時間(*i*-1)と、いま発音開始タイミングとなった音韻*i*、その音高*i*、その発音時間*i*の分析が行われる。ここで、前記音韻*i*の音高*i*は、有声音、無声音を問わず、すべてノートの音高情報KC PITCHに対応する周波数とする。また、前記発音時間*i*は、当該音韻の発音時間設定とメロディーイベントのKEYON、KEYOFF情報などにより決定される。一般に、日本語の歌唱においては、主に先行する子音が所定時間発音され、後続する母音が次のメロディーノートのキーオンまで継続するように前記発音時間*i*を決定する。

【0023】次に、ステップS309に進み、前記声質制御情報記憶部PHONEMEM6のボイスパラメータメモリPMEMから、先行する音韻(*i*-1)後続する音韻*i*に対応する音韻調音結合制御データPHCOMBDATAxを読み出す。次に、ステップS310において、前記ボイスパラメータメモリPMEMから発音を開始すべき音韻*i*に対応する音韻フォルマントデータPHFRMNTDATAyを読み出す。そして、ステップS311に進み、音韻*i*の発音処理中フラグをセットするとともに、音韻*i*の発音開始処理を行う。また、必要に応じて、先行する音韻(*i*-1)の発音の終了処理も行う。このステップS311が終了後この回の歌唱音制御処理を終了する。前記音韻*i*の発音開始処理の詳細については後述する。

【0024】なお、*i*=1のとき、すなわち、当該歌詞の最初の音韻の発音開始処理を行うときには、前記ステップS308における先行する音韻(*i*-1)が存在しない。また、この実施の形態においては、休符や歌唱中に発生する息継ぎの場合に無音区間が発生することとなるが、これらについても音韻として扱うようにしている。したがって、休符や息継ぎの後においては、上記最初の音韻の発音の場合と同様に先行する音韻(*i*-1)が存在しないこととなる。このような場合に対応するために、歌い始めあるいは発音の立上りの遷移状態を示すデータを前記音韻調音結合制御データPHCOMBDATA中に格納しておき、このデータを前記音韻(*i*-1)、音高(*i*-1)として使用するようにしている。

【0025】さて、音韻*i*の発音開始タイミングではなく、前記ステップS307の判断結果がNOとなったときには、ステップS312に進み、音韻*i*の発音処理中

であるか否かが判定される。音韻*i*発音処理中フラグがセットされており、音韻*i*の発音処理中であるときには、この判定結果がYESとなり、ステップS313の発音ボイスパラメータ発生制御処理が行われる。この処理は、前記フォルマント波形発生部81~8mに対して所定時間毎に発音制御パラメータを出力する処理であり、これにより前記フォルマント波形発生部81~8mにより実際に歌唱音が出力される処理である。この処理の詳細については後述する。

【0026】次にステップS314に進み、音韻*i*の発音時間(発音時間*i*)をチェックし、次の音韻(*i*+1)の発音開始タイミングに達したか否かを判定する(ステップS315)。この判定の結果がNOのときは、この回の歌唱音制御処理S50を終了し、前記S60のその他処理に進む。一方、次の音韻(*i*+1)の発音開始タイミングとなったときには、ステップS316に進み、(*i*+1)を前記発音中の音韻シーケンス番号*i*として、前記ステップS308に進む。以下、前記ステップS308~S311を新しい音韻シーケンス番号*i*について実行する。これにより、その音韻*i*の発音の準備処理が行われる。

【0027】図4は、前記S311の音韻*i*発音開始処理を説明するためのフローチャートである。この音韻*i*発音開始処理が開始されると、まず、ステップS401において、この音韻*i*が無音状態から発音される音韻であるか否かが判定される。この判定の結果がYESのときは、直前の音韻の音高がないので、ステップS402に進み、その音韻*i*に対応するメロディーの音高KCPITCH *i*を音韻*i*のピッチデータPITCH *i*および音韻*i*に先行する音韻(*i*-1)のピッチデータPITCH *i*-1にセットする。このときには、音高*i*から音韻*i*が立ち上げられることとなる。なお、このステップS402のようにピッチデータを設定する代わりに、例えば、PITCH *i*-1として、0あるいは他の所定値を設定することもできる。この場合は、その設定された値を初期値として音韻*i*の音高*i*に立ち上がっていくこととなる。

【0028】また、先行して発音される音韻があり、前記ステップS401の判定結果がNOとなったときは、ステップS407に進み、その音韻*i*が無声音であるか否かを判定する。音韻*i*が無声音でありこの判定結果がYESのときは、ステップS408に進み、先行する音韻(*i*-1)のピッチデータPITCH *i*-1が無声音である音韻*i*のピッチデータPITCH *i*とする。これにより、無声音のときに先行する音韻の音高PITCHを保持することができる。また、先行する音韻(*i*-1)が無声音ではないときにはその音韻*i*の音高*i*をそのままPITCH *i*とする。

【0029】さて、上述したようにして音韻*i*および(*i*-1)のピッチデータPITCH *i*およびPITCH *i*-1が設定された後、ステップS403において、PITCH *i*とPIT

CH $i-1$ とが比較される。この結果、 $PITCH\ i > PITCH\ i-1$ であるとき、すなわち、音韻 i の音高 i が先行して発音される音韻 $(i-1)$ の音高 $(i-1)$ よりも高いときには、ステップ S 404 に進む。このステップ S 404 においては、前記ボイスパラメータメモリ PMEM に格納されている音韻フォルマントデータ PHFRMNTDATA 中に含まれている当該音韻 i のピッチの補間レート $Rpitch\ i$ に対し、係数 Ku を乗算して、補間演算において使用するピッチ補間レート $R'pitch\ i$ とする。また、 $PITCH\ i \leq PITCH\ i-1$ でピッチが下降するときには S 403 の判定結果が NO となり、ステップ S 409 において、音韻 i の補間レート $Rpitch\ i$ に対し係数 Kd を乗算して、ピッチ補間レート $R'pitch\ i$ とする。

【0030】ここで、 $0 < Ku < Kd$ とされており、ピッチ上昇時のピッチ補間レートは、ピッチ下降時の補間レートよりも小さく設定されている。なお、前記係数 $Kd = 1$ とし、ピッチの変化無しのおよびピッチが下降するときの補間レートとして、前記音韻フォルマントデータ PHFRMNTDATA 中に含まれている補間レート $Rpitch\ i$ をそのまま用いるようにすればよい。このようにして、ピッチが上昇するときと下降するときとでピッチの遷移レートを変更することができる。

【0031】なお、前記係数 Ku および Kd を、音高差や音高によって異なる値に設定しても良い。例えば、高い音から更に高い音に変化するときには補間レートが異なるようにしてもよい。このようにすることにより、より自然な歌唱音とすることが可能となる。次にステップ S 405 に進み、フォルマント中心周波数の補正演算処理が実行され、次にステップ S 406 において、音韻 i の発音開始指示処理が行われる。この処理は、前記音韻 i 発音処理中フラグをセットする処理である。

【0032】図5は、前記ステップ S 405 の Freq 補正演算処理のフローチャートである。この処理は、発音する音韻が子音 (C) 母音 (V) 子音 (C) であるときに、子音に挟まれて発音される母音のフォルマント中心周波数を制御する処理である。この処理が開始されると、まず、ステップ S 501 において、前記歌詞データをチェックし、先行する音韻 $(i-1)$ 、発音開始タイミングとなった音韻 i および後続する音韻 $(i+1)$ が母音 (V) あるいは子音 (C) のいずれであるかをチェックする。この結果、CVC の連続発音となっている場合には、ステップ S 502 の判定結果が YES となり、ステップ S 503 に進む。このステップ S 503 において、フォルマントの次数を示す変数 j に初期値「1」を代入し、以下、ステップ S 504 ~ S 505、S 507、S 506 および S 509 のループにより、各フォルマント中心周波数の目標値の修正を行う。

【0033】すなわち、まず、ステップ S 504 において、音韻 i の第 j フォルマントの目標中心周波数 $FFreq\ ij$ と先行する音韻 $(i-1)$ の第 j フォルマントの目標

中心周波数 $FFreq\ (i-1)j$ とを比較する。この結果、音韻 i の第 j フォルマントの目標中心周波数 $FFreq\ ij$ が先行する音韻 $(i-1)$ の第 j フォルマントの目標中心周波数 $FFreq\ (i-1)j$ よりも低い周波数であるときには、ステップ S 505 において、音韻 i の第 j フォルマントの目標中心周波数 $FFreq\ ij$ に係数 $Hj1$ ($0 < Hj1 < 1$) を乗算して、低く設定された新たな第 j フォルマントの目標中心周波数 $FFreq'\ ij$ とする。

【0034】一方、音韻 i の第 j フォルマントの目標中心周波数 $FFreq\ ij$ が先行する音韻 $(i-1)$ の第 j フォルマントの目標中心周波数 $FFreq\ (i-1)j$ よりも高い周波数であるときには、ステップ S 507 において、音韻 i の第 j フォルマントの目標中心周波数 $FFreq\ ij$ に係数 $Hj2$ ($1 \leq Hj2$) を乗算して、高く設定された新たな第 j フォルマントの目標中心周波数 $FFreq'\ ij$ とする。なお、音韻 i の第 j フォルマントの目標中心周波数 $FFreq\ ij$ と先行する音韻 $(i-1)$ の第 j フォルマントの目標中心周波数 $FFreq\ (i-1)j$ とが一致しているときは、前記 $Hj2 = 1$ とし、目標中心周波数の修正を行わないようにする。

【0035】次に、ステップ S 506 に進み、すべてのフォルマントについての処理が終了した否かを判定し、終了していないときは、変数 j をインクリメントして (S 509)、前記ステップ S 504 からの処理を繰り返す。また、すべてのフォルマントについての処理が終了したときには、この Freq 補正演算処理を終了する。

【0036】また、前記ステップ S 502 の判定が NO となったとき、すなわち、CVC の連続発音となっていないときには、フォルマントの目標中心周波数を変更する必要がないのであるから、音韻 i に対応する各フォルマントの目標中心周波数 $[FFreq]\ i$ をそのまま目標中心周波数 $[FFreq']\ i$ として、この Freq 補正演算処理を終了する。以上が、前記ステップ S 311 の音韻 i 発音開始処理の内容である。このようにして、前記図3において音韻 i の発音開始タイミングとなったときの発音準備処理が終了する。

【0037】次に、前記ステップ S 313 の発音ボイスパラメータ発生制御処理について、図6のフローチャートを参照して説明する。前述したように、このステップ S 313 は音韻 i の発音処理中に実行される処理であり、前記フォルマント波形発生部 81 ~ 8m に対し所定時間毎に各パラメータを供給して、所定の音韻を発生させる処理である。

【0038】この処理においては、まず、ステップ S 601 において、先行する音韻 $(i-1)$ が有声音であるか否かが判定される。先行する音韻 $(i-1)$ が有声音の場合にはこの判定結果が YES となり、次に、ステップ S 602 において発音処理中の音韻 i が有声音であるか否かが判定される。また、先行する音韻 $(i-1)$ が無声音のときにはステップ S 606 に進み、当該音韻 i

が無声音であるか否かが判定される。

【0039】そして、前記ステップS602の判定結果がYESのとき、すなわち、先行する音韻(i-1)が有声音で当該音韻iが有声音であり有声音の発音が連続するとき、および、前記ステップS606の判定結果がNOのとき、すなわち、先行する音韻(i-1)が無声音で当該音韻iが有声音のときには、ステップS603の処理が実行される。このステップS603は、先行する音韻(i-1)のピッチPITCH i-1と現在発音している音韻iのピッチPITCH iとの間を前記ステップS404あるいはS409において設定した補間レートR'pitchで補間し、その結果を所定時間(例えば数msec)毎に、前記フォルマント波形発生部81~8mにPITCHデータとして出力する処理である。これにより、各フォルマント波形発生部81~8mにおいて生成される音韻のピッチが設定されたレートで変更される。

【0040】また、前記ステップS602の判定結果がNOのとき、すなわち、先行する音韻(i-1)が有声音で後続する音韻iが無声音のとき、および、前記ステップS606の判定結果がYESのとき、すなわち、先行する音韻(i-1)が無声音で後続する音韻iが無声音のときは、ステップS607が実行される。このステップS607においては、その時点のピッチPITCHを保持したまま、すなわち、補間演算処理を行うことなく、所定時間毎にPITCHデータとして、前記フォルマント波形発生部81~8mへ出力する。これにより、無声音のときはピッチPITCHが保持されることとなる。

【0041】上記ステップS603あるいはS604においてピッチデータの送出を実行した後は、ステップS604に進み、各フォルマント周波数の制御データの送出が行われる。この処理においては、先行する音韻(i-1)と当該音韻iとの間の音韻調音結合制御データPHCOMBDATAxiに応じた調音結合特性で、各フォルマント周波数について、[FFreq] i-1~[FFreq'] iの間で補間演算し、所定時間毎に補間値を出力する。ここで、[FFreq] i-1は先行する音韻(i-1)の各フォルマントの中心周波数であり、[FFreq'] iは、前記FFreq補正演算処理S405において演算された当該音韻iの各フォルマントの目標中心周波数である。

【0042】次に、ステップS605に進み、他の発音ボイスパラメータ、FShape、FLevel等についても、同様に補間演算処理を行い、所定時間毎に、前記フォルマント波形発生部81~8mに出力する。このようにして、所定時間(例えば、数msec)毎に、各フォルマント波形発生部81~8mに発音ボイスパラメータが送出され、各フォルマント波形発生部81~8mにおいて前述のようにして当該音韻の対応するフォルマント波形が生成される。各フォルマント波形発生部81~8mから出力される各フォルマントに対応する出力は信号合成部11において加算され、当該歌詞に対応した合成された音

韻が発音されることとなる。

【0043】なお、以上の説明においては、フォルマント中心周波数、フォルマントレベル、フォルマント帯域幅およびピッチ周波数などの各パラメータを、前記制御部1から所定時間間隔で(例えば、数ミリ秒程度の間隔で)逐次送出して制御するようにしていたが、この時間間隔をより長くして、各フォルマント波形発生部81~8mに含まれているエンベロープジェネレータにより前記各パラメータを逐次制御させるようにしてもよい。

【0044】また、上記においては、前記声質制御情報記憶部PHONEMEM6中に複数種類の音韻フォルマントデータPHFRMNTDATAおよび音韻調音結合制御データPHCOMBDATAを記憶し、発生させたい声質に応じてそれらを選択するようにしていたが、記憶されている複数種類のデータ(例えば、男声、女声あるいは個人の声を分析して得た各個人に対応するデータ)のうちのいくつかを選択して、それらを任意に組み合わせ、補間処理を行うことにより、それらの中間的な特性あるいは新規な特性を有する音韻フォルマントデータあるいは音韻調音結合制御データを生成し、そのデータを用いて発音制御するようにすることもできる。例えば、男声と女声の2種類のデータから音韻フォルマントデータを生成し、該データを発音制御に用いることにより、男女2者の中間的な声質を持った音韻を発生させることが可能となる。

【0045】本発明の歌唱音合成装置の適用分野として特に好適な例を挙げれば、歌唱音出力可能な電子楽器やコンピュータシステム、音声応答装置、あるいはゲームマシンやカラオケなどのアミューズメント機器などが考えられる。また、本発明の歌唱音合成装置は、パソコンに代表されるコンピュータシステムのソフトウェアという形態で実施することも可能である。その際、音声波形合成までCPUにより実行するようにしてもよいし、あるいは図1に示したように別途音源を設けてもよい。

【0046】

【発明の効果】以上説明したように、本発明の歌唱音合成装置によれば、音程が変化するときにおけるピッチの変化速度を音程の上昇時は遅くし下降時は早くしているために、音程が変化するときにおいても自然な歌唱音を発音することができる。また、音程が変化するときにおけるピッチのつながりがスムーズになり、自然な発音となる。さらに、子音、母音、子音の順に発音させるときに、母音のフォルマント中心周波数を、単独で発音させる場合のフォルマント中心周波数に達しないうちに後続の子音のフォルマント中心周波数に戻すようにしているために、違和感のない歌唱音を出力することが可能となる。

【図面の簡単な説明】

【図1】 本発明の歌唱音合成装置のシステム構成の一例を示すブロック図である。

【図2】 本発明の歌唱音合成装置におけるメインルー

チンを示すフローチャートである。

【図3】 歌唱音制御処理を説明するためのフローチャートである。

【図4】 音韻 i 発音開始処理を説明するためのフローチャートである。

【図5】 FFReq補正演算処理を説明するためのフローチャートである。

【図6】 発音ボイスパラメータ発生制御処理を説明するためのフローチャートである。

【図7】 音高の上昇、下降時の状態を説明するための図である。

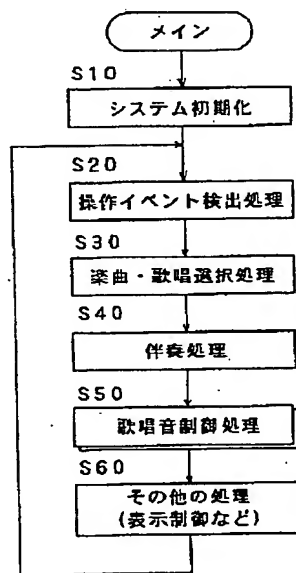
【図8】 音高が変化したときの無声音と有声音出力を説明するための図である。

【図9】 子音母音子音の発音時のフォルマント周波数の変動を説明するための図である。

【符号の説明】

1 制御部、2 プログラム記憶装置、3 操作設定部、4 MIDIインターフェース、5 楽曲情報記憶部、6 声質制御情報記憶部、7 発音ボイスパラメータ供給制御部、8 1～8m フォルマント波形発生部、9 無声音音源部、10 有声音音源部、11 信号合成部、12 伴奏用音源部

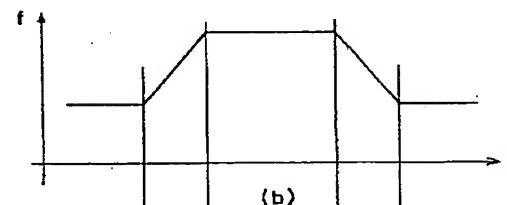
【図2】



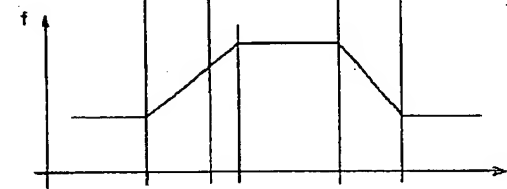
【図7】



(a)

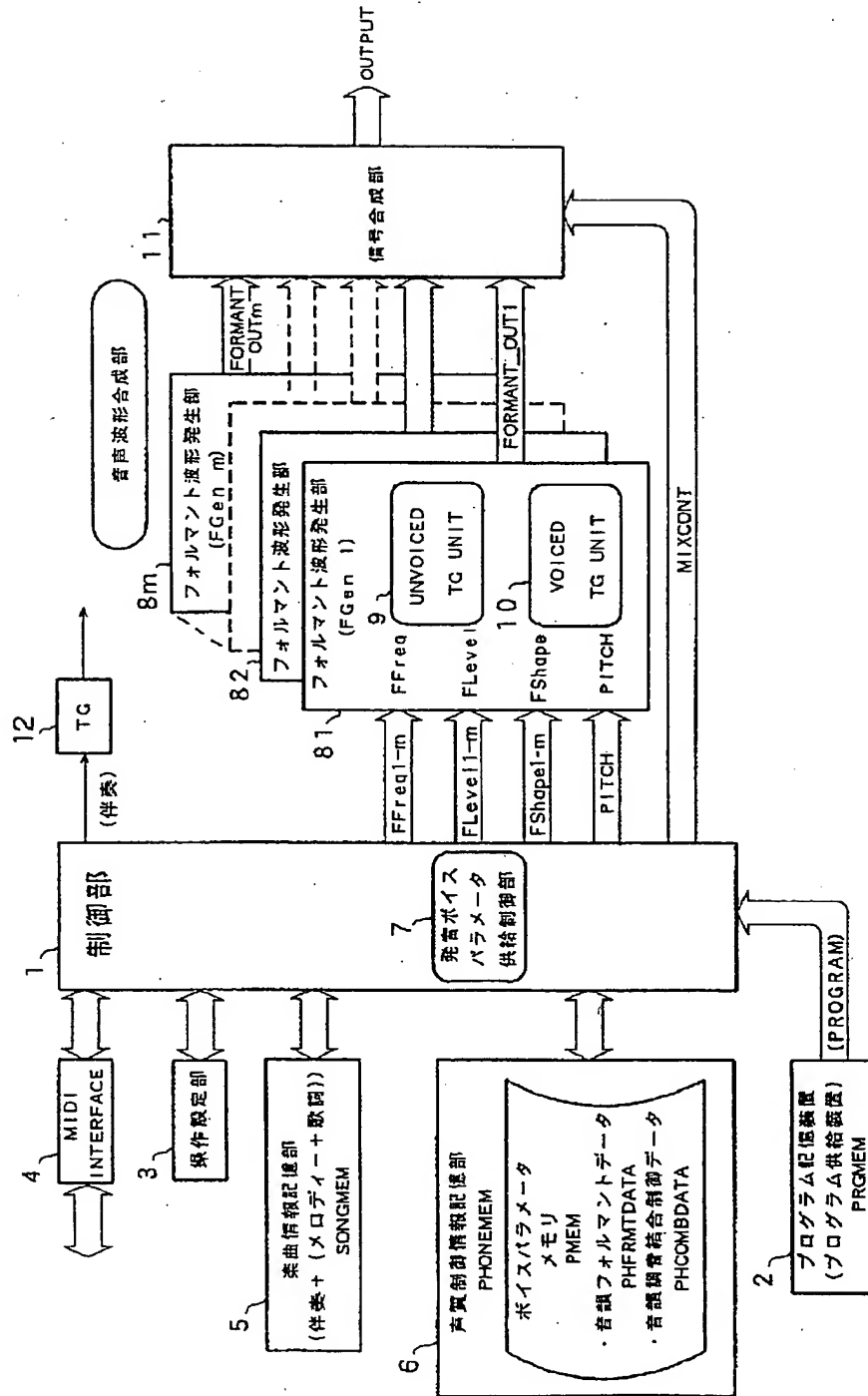


(b)

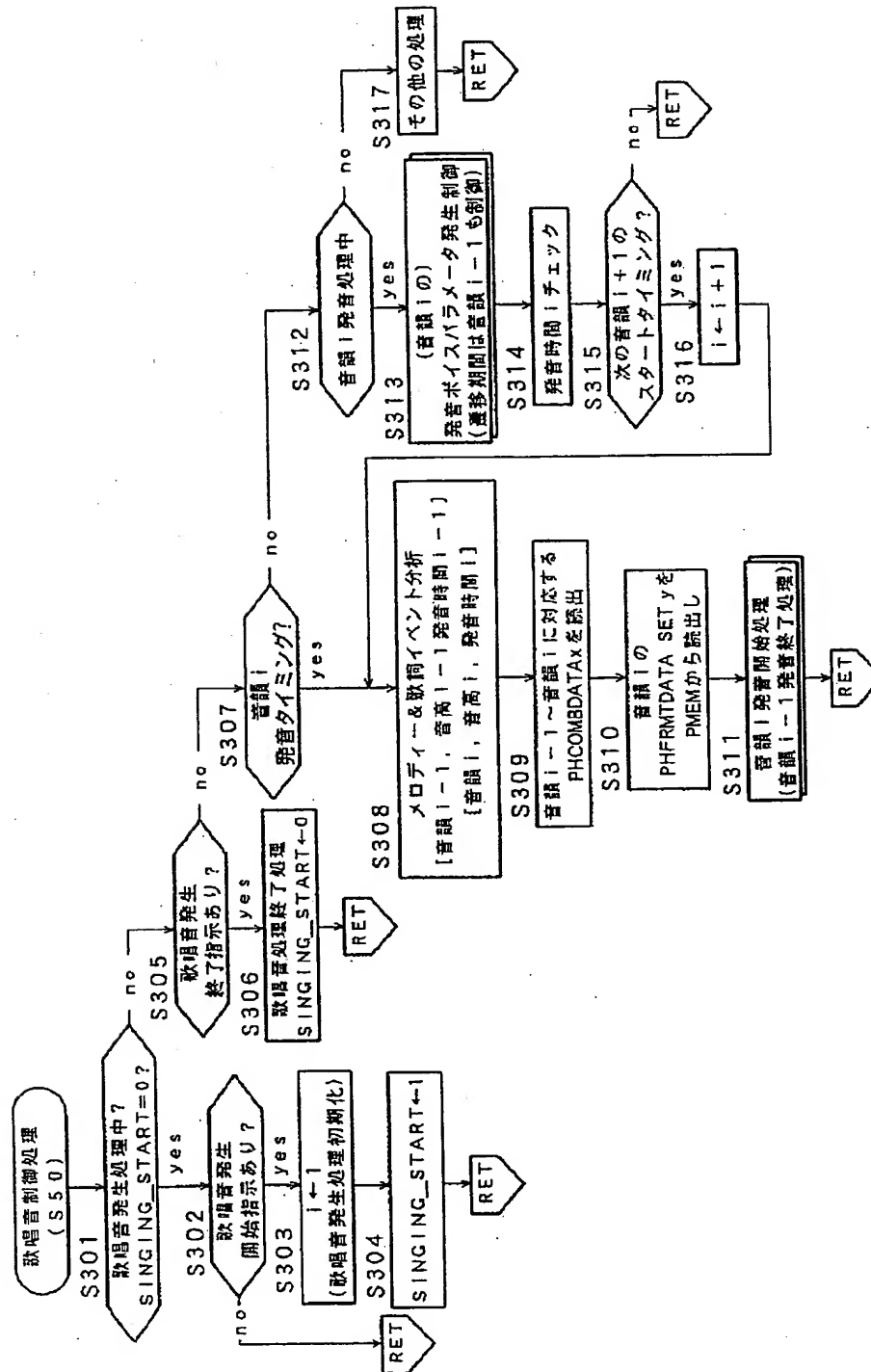


(c)

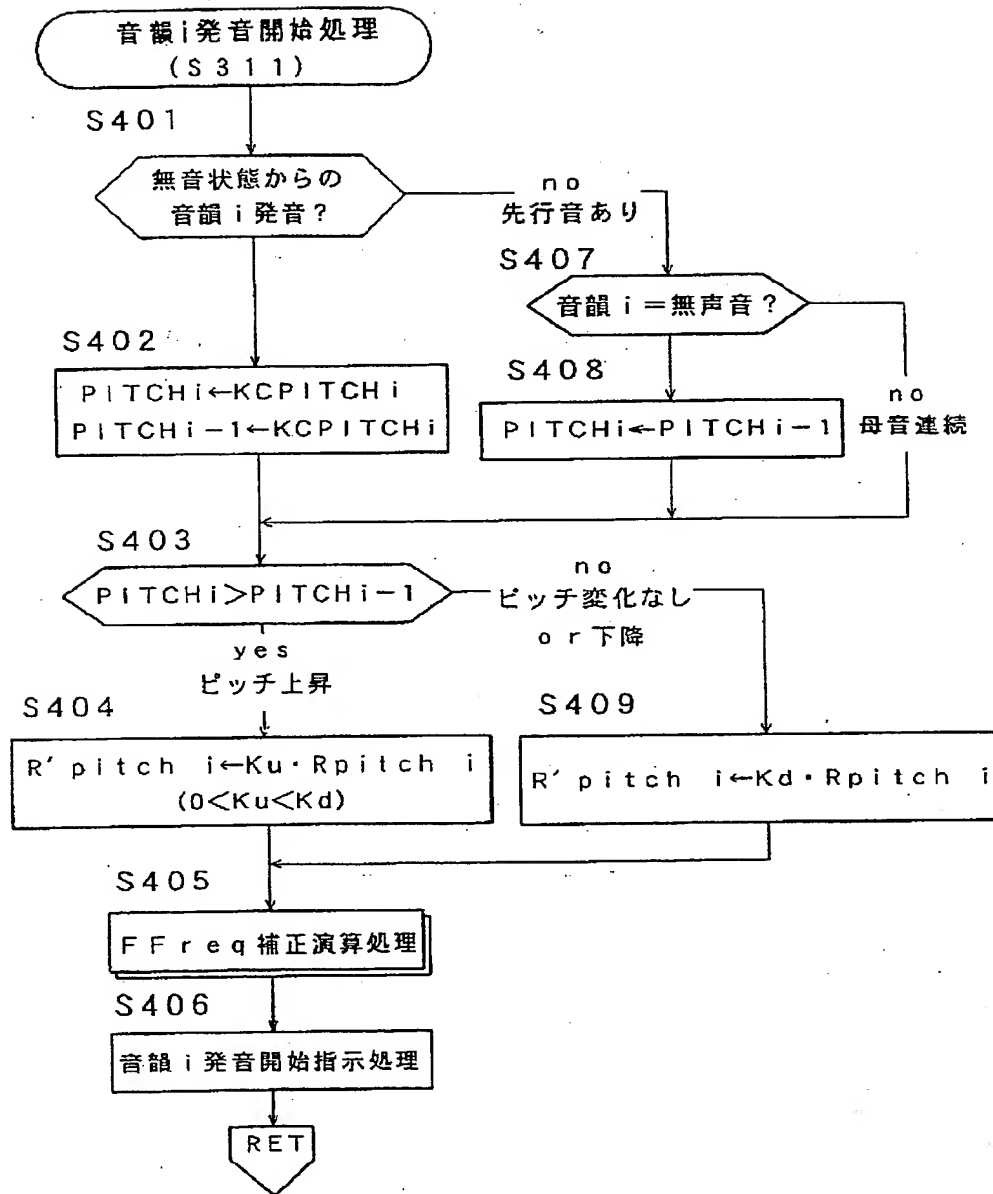
【図 1】



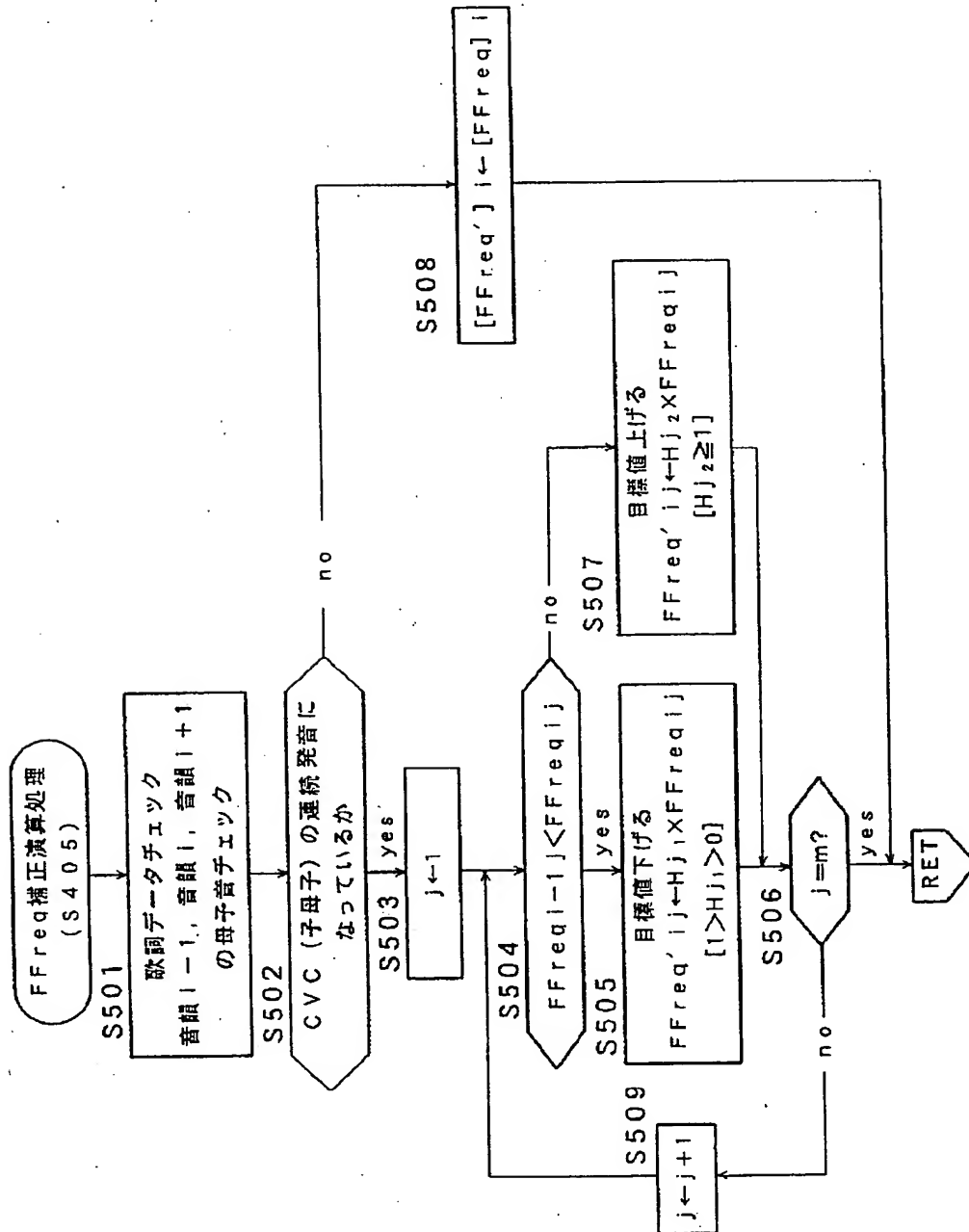
【図 3】



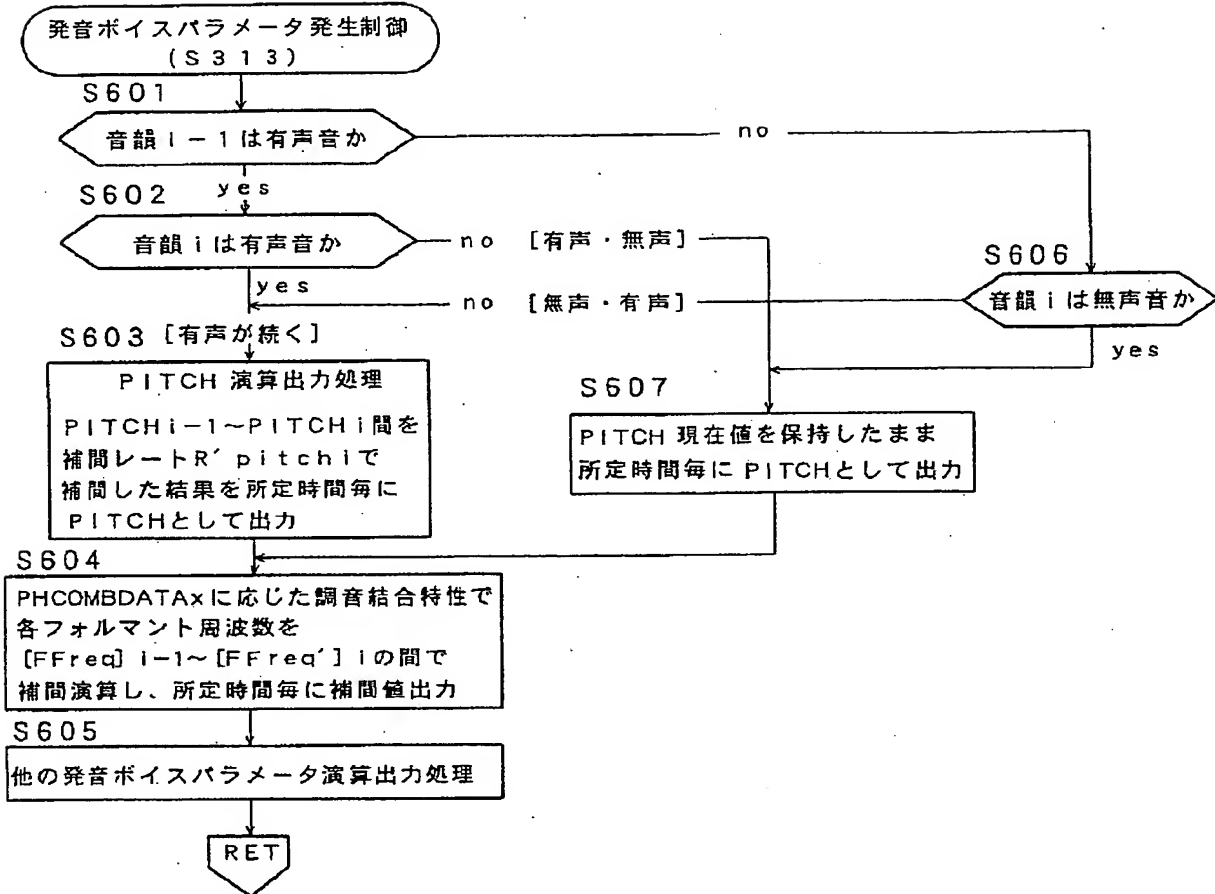
【図 4】



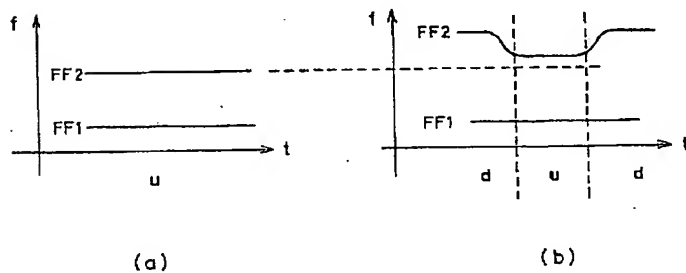
【図 5】



【図6】



【図9】



【図 8】

